included in X_2 in his model. Nelson used the ML method in his work. The likelihood function for this model is

$$L(\beta_1,\beta_2,\sigma_1^2,\sigma_2^2,\sigma_{12}) = \prod_i \left(\int_{-\infty}^{g_{2i}} (g_{1i}, u_{2i})\, du_{2i}\right)^{I_i} \left(1-\Phi_i\right)^{1-I_i} \tag{8.27}$$

where

$$g_{1i} = y_i - \beta_1' X_{1i}$$
$$g_{2i} = y_i - \beta_2' X_{2i}$$
$$\Phi_i = \Phi\left(\frac{\beta_1' X_{1i} - \beta_2' X_{2i}}{\sigma}\right)$$

Consider next the bank-borrowing model. Here the estimation of the probit model and the parameter estimates it yields is the same as before. As for equation (8.21), we can now estimate it by OLS, because we have all observations on y_1. It can be easily verified that the conditions for identification are the same as those in the Gronau–Nelson model. The likelihood function for this model is

$$L(\beta_1,\beta_2,\sigma_1^2,\sigma_2^2,\sigma_{12}) = \prod_i \left[\frac{1}{\sigma_1}\exp\left(-\frac{1}{2\sigma_1^2}(y_{1i}-\beta_1' X_{1i})^2\right)\right]\left(\Phi_i\right)^{I_i}\left(1-\Phi_i\right)^{1-I_i} \tag{8.28}$$

where Φ_i is defined in (8.27).

Note that the conditions for identification in the switching regression model considered in the preceding section will be the same as those for the censored regression model considered here if there are not enough observations in the second regime to be able to estimate the parameters in (8.13).

Finally, note that the two-stage estimation procedure can be applied using the observations on all the individuals – working and nonworking. The procedure is the same as that suggested for the tobit model in equation (8.4). We note that

$$E(y_{1i} \mid I_i = 1) = \beta_1' X_{1i} - \sigma_{1u}\frac{\phi_i}{\Phi_i}$$

Hence, the unconditional expectation of y_{1i} is

$$E(y_{1i}) = E(y_{1i} \mid I_i=1)P(I_i=1) + E(y_{1i} \mid I_i=0)P(I_i=0)$$
$$= \beta_1' \Phi_i X_{1i} - \sigma_{1u}\phi_i$$

We estimate this equation by the two-stage method, substituting $\hat{\Phi}_i$ and $\hat{\phi}_i$ for Φ_i and ϕ_i, respectively, instead of equation (8.12). In this procedure we use the observations on all individuals, not just the working ones as in the estimation of (8.12).

8.5 Two-stage estimation of Heckman's model

Heckman (1974) considered a model of labor supply in which wages and hours worked are the two endogenous variables. The model consists of the shadow-wage equation

$$S = \gamma_0 + \gamma_1 H + \gamma_2 Z + u_1 \tag{8.29}$$

and the market-wage equation

$$W = \beta_0 + \beta_1 X + u_2 \tag{8.30}$$

where X and Z are exogenous variables. Heckman assumed that hours worked H adjust, so that $S=W$. Hence, from equations (8.29) and (8.30), we get

$$H = \frac{\beta_0+\beta_1 X-\gamma_0-\gamma_2 Z}{\gamma_1} + \frac{u_2-u_1}{\gamma_1} \tag{8.31}$$

If $H>0$, the person is in the labor force, and we observe H and W. If $H \leqslant 0$, the person is not in the labor force. For the observations for which $H \leqslant 0$, we have

$$\frac{u_2-u_1}{\gamma_1} < \frac{\gamma_0-\beta_0+\gamma_2 Z-\beta_1 X}{\gamma_1}$$

or

$$u_2 - u_1 < \gamma_0 - \beta_0 + \gamma_2 Z - \beta_1 X$$

because γ_1 is expected to be positive. If $\mathrm{Var}(u_2-u_1)=\sigma^2$, then

$$\mathrm{Prob}(H \leqslant 0) = \Phi\left(\frac{\gamma_0-\beta_0+\gamma_2 Z-\beta_1 X}{\sigma}\right) \tag{8.32}$$

where $\Phi(\cdot)$ is the distribution function of the standard normal. Thus, the likelihood function for this model is

$$L = \prod_{H>0} F(W,H) \cdot \prod_{H\leqslant 0} \Phi(\Delta) \tag{8.33}$$

where

$$\Delta = \frac{\gamma_0-\beta_0+\gamma_2 Z-\beta_1 X}{\sigma} \tag{8.34}$$

Heckman (1974) estimated this model by ML methods. Later (Heckman, 1976*b*), he suggested a two-stage estimation method.[4] For the two-stage method we need to evaluate $E(u_2 \mid H>0)$ in equation (8.30). This is

[4] The discussion of the two-stage method that follows is not exactly the same as that exposited by Heckman. We shall follow the line of reasoning in the preceding sections.

$$E\left(u_2 \middle| \frac{u_2-u_1}{\sigma} > \Delta\right)$$

Denoting $\text{Var}(u_2)=\sigma_2^2$ and $\text{Cov}(u_2,u_1)=\sigma_{12}$, we get this [see equation (8.10)] as

$$\frac{\sigma_2^2-\sigma_{12}}{\sigma} \frac{\phi(\Delta)}{1-\Phi(\Delta)}$$

Thus, the wage equation (8.30) can be written as

$$W=\beta_0+\beta_1 X+\frac{\sigma_2^2-\sigma_{12}}{\sigma} \frac{\phi(\Delta)}{1-\Phi(\Delta)}+V_2 \tag{8.35}$$

where V_2 is the new residual with the property that $E(V_2)=0$.

The two-stage procedure that Heckman suggested is to get estimates of the parameters in Δ from a probit ML estimation and then estimate equation (8.35) by OLS after substituting $\phi(\hat{\Delta})/[1-\Phi(\hat{\Delta})]$ for $\phi(\Delta)/[1-\Phi(\Delta)]$. This gives us consistent estimates of β_0, β_1, and $(\sigma_2^2-\sigma_{12})/\sigma$. To see how we can get estimates of all the parameters, note that the probit ML gives estimates of $(\gamma_0-\beta_0)/\sigma$:

$\dfrac{\gamma_{2j}-\beta_{1j}}{\sigma}$ for the common variables in X and Z

$\dfrac{\gamma_{2j}}{\sigma}$ for the variables in Z not in X

$\dfrac{\beta_{1j}}{\sigma}$ for the variables in X not in Z

Now, examining equations (8.29) and (8.30), we note that (if we assume $\sigma_{12}\neq 0$) the condition for identification is that there be at least one variable in X not included in Z. For this excluded variable we get an estimate of β_{1j}/σ from the probit equation, and because we have an estimate of β_{1j} from the wage equation, we can get an estimate of σ. Once we have an estimate of σ and estimates of β_0 and β_1, it is easy to see that all elements of γ_0 and γ_2 can now be estimated. Also, because $\sigma^2=\sigma_1^2+\sigma_2^2-2\sigma_{12}$, we have an estimate of $\sigma_1^2-\sigma_{12}$ (from the estimates of σ^2 and $\sigma_2^2-\sigma_{12}$). The problem now is to get separate estimates of σ_1^2, σ_2^2, and σ_{12}. For this purpose we have to use the estimated residuals in the wage equation, as we did in the preceding section for the switching regression model.

Note that we have, in a fashion analogous to equation (8.15),

$$E(u_{2i}^2 \mid H>0)=E\left(u_{2i}^2 \middle| \frac{u_{2i}-u_{1i}}{\sigma} > \Delta_i\right)$$

$$=\sigma_2^2+\left(\frac{\sigma_2^2-\sigma_{12}}{\sigma}\right)^2 \Delta_i \frac{\phi(\Delta_i)}{1-\Phi(\Delta_i)} \tag{8.36}$$

Thus, after computing the residuals

$$\hat{u}_{2i}=W_i-\hat{\beta}_0-\hat{\beta}_1 X_i \tag{8.37}$$

we estimate σ_2^2 by

$$\hat{\sigma}_2^2=\frac{1}{N_1}\sum_{i=1}^{N_1}\left[\hat{u}_{2i}^2-\left(\frac{\sigma_2^2-\sigma_{12}}{\sigma}\right)^2 \hat{\Delta}_i \frac{\phi(\hat{\Delta}_i)}{1-\Phi(\hat{\Delta}_i)}\right] \tag{8.38}$$

where N_1 is the number of observations for which $H>0$. Once we obtain an estimate of σ_2^2, we have estimates of σ_{12} and σ_1^2, because we have earlier estimated $\sigma_2^2-\sigma_{12}$ and $\sigma_1^2-\sigma_{12}$. Finally, to obtain an estimate of γ_1, we have to estimate equation (8.31) by the tobit method.

The essential features of the two-stage methods are clear. First we obtain the expected values of the residuals that are truncated. These expected values involve unknown parameters, but these usually can be estimated by the probit method. We then introduce the estimated values of these variables into the original equation and estimate it by ordinary least squares (or weighted least squares if we take account of the heteroscedasticity problems). Note that to obtain estimates of the residual variances we have to compute the estimated residuals and use formulas like (8.17) and (8.18).

In the foregoing discussion of Heckman's model, we assumed $\sigma_{12}\neq 0$. If $\sigma_{12}=0$, then we do not need the condition (for identification) that there be at least one variable in X not included in Z. But now the procedure for getting consistent estimates of the parameters is different. From two-stage estimation of the wage equation (8.35) we get estimates of β_0, β_1, and σ_2^2/σ. We next estimate σ_2^2 from the computed residuals for the wage equation according to formula (8.38), using $\sigma_{12}=0$. We can thus get an estimate of σ, and now the estimates from the probit ML lead us to estimates of γ_0 and γ_2. Because $\sigma^2=\sigma_1^2+\sigma_2^2$, we also have an estimate of σ_1^2. As before, γ_1 has to be estimated by using the tobit method for equation (8.31).

Note that the Heckman model can be written as

$$\left.\begin{aligned} y_1&=\gamma_0+\gamma_1 y_2+\gamma_2 Z+u_1 \\ y_2&=\beta_0+\beta_1 X+u_2 \end{aligned}\right\} \text{ if } y_2>0$$
$$y_1=y_2=0 \quad \text{otherwise} \tag{8.39}$$

The conditions for identification for the simultaneous-equations model (8.39) are well known; namely, $\text{Cov}(u_1,u_2)=0$, or there is at least one variable in X not included in Z. These are the conditions for identification in Heckman's model.

Finally, note that the two-stage estimation of the wage equation, as described in equation (8.35), uses only the working individuals. One can easily extend this to cover data on all individuals (working and nonworking), along the lines described at the end of the preceding section.

8.6 Two-stage estimation of structural equations

What we discussed in the preceding section was probit and tobit estimation of the hours-worked equation and two-stage estimation of the wage equation (which is anyhow in its reduced form). Thus, both the equations being estimated are reduced-form equations, and we discussed how to recover the structural parameters from the estimates of these reduced-form parameters. One can, however, think of estimating the structural equation (8.29) directly by two-stage methods to obtain estimates of γ_0, γ_1, and γ_2. To do this, note that $S=W$ and

$$E(u_1 \mid H>0) = E\left(u_1 \,\middle|\, \frac{u_2-u_1}{\sigma} > \Delta\right) = \frac{\sigma_{12}-\sigma_1^2}{\sigma} \frac{\phi(\Delta)}{1-\Phi(\Delta)}$$

where Δ is defined in (8.34). We can now write equation (8.29) as

$$W = \gamma_0 + \gamma_1 H + \gamma_2 Z + \frac{\sigma_{12}-\sigma_1^2}{\sigma} \frac{\phi(\Delta)}{1-\Phi(\Delta)} + v_1 \tag{8.40}$$

where v_1 is the residual u_1 corrected for its mean, and hence $E(v_i)=0$. However, we cannot estimate equation (8.40) by OLS, because H is an endogenous variable. What we have to do is get an estimate of H for the subsample $H>0$. Note that

$$E\left(\frac{u_2-u_1}{\gamma_1} \,\middle|\, H>0\right) = \frac{\sigma}{\gamma_1} E\left(\frac{u_2-u_1}{\sigma} > \Delta\right) = \frac{\sigma}{\gamma_1} \frac{\phi(\Delta)}{1-\Phi(\Delta)}$$

Hence, equation (8.31) for $H>0$ can be written as

$$H = \frac{\beta_0-\gamma_0}{\gamma_1} + \frac{\beta_1 X-\gamma_1 Z}{\gamma_1} + \frac{\sigma}{\gamma_1} \frac{\phi(\Delta)}{1-\Phi(\Delta)} + v \tag{8.41}$$

where $E(v)=0$. We estimate this equation by OLS after substituting $\hat{\Delta}$ for Δ. After estimating the parameters in (8.41) by this two-stage method, we get the estimated values $\hat{H}$. Note that these are obtained from (8.41), not (8.31). We substitute these values of $\hat{H}$ in place of H in equation (8.40) and estimate that equation by OLS. Alternatively, we can use $\hat{H}$ as an instrumental variable in estimating (8.40). This will also produce consistent estimates.

It can be easily shown that the resulting two-stage estimates of the structural parameters are consistent. Their asymptotic covariance matrix can be derived by the methods in the Appendix at the end of this chapter. The method suggested here is the appropriate analogue of the usual two-stage least-squares method for the case of truncated variables. Briefly, what we do is (a) evaluate the expected values of the residuals in both the structural and reduced-form equations, (b) estimate the reduced-form equations by the two-stage estimation method, and (c) substitute the estimated values of the endogenous variables on the right-hand side of each structural equation, as obtained from step (b), and use OLS to estimate the parameters of the structural equations.

The main difference between the ordinary two-stage least-squares (2SLS) procedure and the one suggested here lies in noting the fact that the residuals in the structural equations and the reduced-form equations do not have zero means and that the estimates from the reduced forms (to be substituted in the structural equations) should take this into account.

The advantage of the procedure described here, which was suggested by Lee et al. (1980), is that it can be used in any simultaneous-equations model with truncation. For instance, suppose that hours worked also occurs as an explanatory variable in the market-wage equation (8.20); then this equation also is in a structural form (not a reduced form as in Heckman's model). Again, one can proceed via the reduced-form estimation, as outlined in the preceding section. But this creates problems of multiple solutions for the structural parameters if the equations are overidentified. Hence, it will be desirable to estimate the structural equations by the two-stage methods. Equation (8.35) will now be changed to

$$W = \beta_0 + \beta_1 H + \beta_2 X + \frac{\sigma_2-\sigma_{12}}{\sigma} \frac{\phi(\Delta)}{1-\Phi(\Delta)} + v_2 \tag{8.42}$$

and we estimate this equation by OLS after substituting $\hat{H}$ for H, as obtained from a two-stage estimation of (8.41), and $\hat{\Delta}$ for Δ, as obtained from the probit ML method.

The procedure described is very general and can be used in the estimation of all simultaneous-equations systems with censored dependent variables. The procedure gets complicated, however, if there is more than one condition to determine the censoring. In the Heckman model, for instance, there is only one condition: $H>0$. Amemiya (1974*b*) considered a model in which there are two censoring conditions involved:

$$y_1 = \gamma_1 y_2 + \beta_1' X_1 + u_1 \quad \text{if } y_1 > 0$$
$$y_1 = 0 \quad \text{otherwise}$$

$$y_2 = \gamma_2 y_1 + \beta_2' X_2 + u_2 \quad \text{if } y_2 > 0$$
$$y_2 = 0 \quad \text{otherwise}$$

Amemiya showed that this model is logically consistent only if $\gamma_1\gamma_2 < 1$. We can divide the observations into the following sets:

$$S_1: \; y_1 > 0, \quad y_2 > 0$$
$$S_2: \; y_1 > 0, \quad y_2 = 0$$
$$S_3: \; y_1 = 0, \quad y_2 > 0$$
$$S_4: \; y_1 = 0, \quad y_2 = 0$$

Consider, as Amemiya did, only observations in S_1. To apply the two-stage methods described earlier, we need to evaluate the expectations of the residuals in the structural equations and the reduced forms under the condition $y_1 > 0$, $y_2 > 0$. These obviously involve double integrals. The first stage of the two-stage method in this case involves estimation of a bivariate probit equation. The rest of the steps, of course, are straightforward. In this model, as with two-stage estimation of the simple tobit model, there is not much to be gained in using the two-stage method as compared with the maximum-likelihood method.

There is another case in which two-stage estimation involves substitution of estimated endogenous variables in an equation prior to estimation. This is the case of the switching regression model considered earlier in equations (8.5) and (8.6), where the criterion function involves the values of y_1 in the two regimes. Specifically, the model in (8.5) and (8.6) is changed as follows:

$$\text{Regime 1:} \quad y_{1i} = \beta_1' X_{1i} + u_{1i} \tag{8.43}$$

$$\text{Regime 2:} \quad y_{2i} = \beta_2' X_{2i} + u_{2i} \tag{8.44}$$

and

$$C_i = \gamma' Z_i + \delta(y_{1i} - y_{2i}) - u_i \tag{8.45}$$

Only one of y_{1i} or y_{2i} is observed for each individual i, depending on whether $C_i \geqslant 0$ or $C_i < 0$. The criterion function (8.45), however, involves $y_{1i} - y_{2i}$, and to estimate δ we need to estimate y_{1i} and y_{2i} for all the observations. Examples of this model are the following:

1. The union model by Lee (1978), where y_{1i} are wages in the union sector, y_{2i} are wages in the nonunion sector, and the criterion function that determines whether or not an individual joins the union depends on, in addition to other factors, the expected benefit as measured by $y_1 - y_2$. Interest centers on whether or not $y_1 - y_2$ is a significant variable in the decision function, that is, whether or not the coefficient of δ in (8.45) is significant.
2. The college-education model by Willis and Rosen (1979), where y_{1i} are earnings of college graduates, y_{2i} are earnings of those who are not college graduates, and the criterion function that determines whether or not an individual goes to college depends on, in addition to other factors, the expected benefit as measured by $y_1 - y_2$. Interest again centers on whether or not this variable is a significant variable in the decision function, that is, whether or not δ is significant.

Estimation of this model proceeds along the lines described earlier. We first write the criterion function in its reduced form:

$$C_i = \gamma' Z_i + \delta(\beta_1' X_{1i} - \beta_2' X_{2i}) + \delta(u_{1i} - u_{2i}) - u_i \tag{8.46}$$

We can write this as

$$C_i = \gamma^{*\prime} Z_i^* - u_i^* \tag{8.47}$$

where γ^* is defined suitably. Now the two-stage estimation of the two wage equations proceeds as before. We define

$$I_i = 1 \quad \textit{if } C_i > 0$$
$$I_i = 0 \quad \text{otherwise}$$

Based on the observations on I_i, we use the probit method to get an estimate of γ^*.

Next we estimate equations of the form (8.12) and (8.13) (adding the superscript asterisk where applicable) and get estimates of the parameters β_1 and β_2. Everything is as before except the estimation of the parameters γ and δ in the criterion function (8.45). (Note that these parameters are estimable only up to a proportionality factor.) For this, we obtain predicted values of y_{1i} and y_{2i} for all observations as

$$\hat{y}_{1i} = \hat{\beta}_1' X_{1i} \quad \text{and} \quad \hat{y}_{2i} = \hat{\beta}_2' X_{2i} \tag{8.48}$$

Note that we do not use equations (8.12) and (8.13) as we did in the two-stage estimation of structural equations described earlier. There the reason these equations were used, and not (8.48), was that only a subset of the observations was used. We next estimate δ by using the probit ML method applied to equation (8.45) based on the dichotomous observations I_i. This method is called the *structural probit method* (it is the use of probit method to estimate a structural equation after substituting the estimates of the endogenous variables in the equation). Lee (1979*a*) showed that the resulting estimates of γ and δ are consistent and derived the asymptotic covariance matrix.

The question of which estimates of the endogenous variables have to model). Another major finding is that most new entrants prefer union jobs but cannot get them. As time goes by, and they accrue nonunion seniority, they become less likely to want union jobs. Thus, the union status of most workers is determined by their success in being selected from the queue early in their working lives.

Abowd and Farber used only the decision functions, not the wage equations, in their analysis. This latter model would be an interesting one to estimate. In Chapter 9 we discussed some of these models.

11.7 Summary and conclusions

The purpose of this chapter has been to review several methods discussed in the previous chapters in relation to a single problem and point out how some mechanical formulations endogenizing dummy variables result in models that are not entirely satisfactory. The proper way to model is to incorporate the choice processes explicitly.

Appendix: Some results on truncated distributions

The truncated normal distribution

We shall use the symbol ϕ to denote the density function and Φ to denote the cumulative distribution function of the standard normal distribution. We shall also use the notation f and F to denote the density function and distribution function for distributions other than the standard normal.

$$\Phi(y) = \int_{-\infty}^{y} \frac{1}{(2\pi)^{1/2}} \exp(-\tfrac{1}{2}u^2)\, du$$

and

$$\phi(y) = \frac{d\Phi(y)}{dy} = \frac{1}{(2\pi)^{1/2}} \exp(-\tfrac{1}{2}y^2)$$

For a normal distribution with mean μ and variance σ^2, the function $\Phi(y)$ becomes $\Phi[(y-\mu)/\sigma]$, and the function $\phi(y)$ becomes $(1/\sigma)\phi[(y-\mu)/\sigma]$.

Mean and variance of the truncated normal

Suppose the random variable X is $N(0,1)$, and we consider the truncated distribution $X \geqslant c_1$. The mean and variance of this truncated distribution are given by

$$E(X) = \frac{\phi(c_1)}{1-\Phi(c_1)} = \frac{\text{ordinate at } X=c_1}{\text{right-hand tail area}} = M_1 \quad \text{(say)}$$

$$V(X) = 1 - M_1(M_1 - c_1)$$

If the truncation is from above, so that we consider the distribution $X \leqslant c_2$, then

$$E(X) = \frac{-\phi(c_2)}{\Phi(c_2)} = M_2 \quad \text{(say)}$$

$$V(X) = 1 - M_2(M_2 - c_2)$$

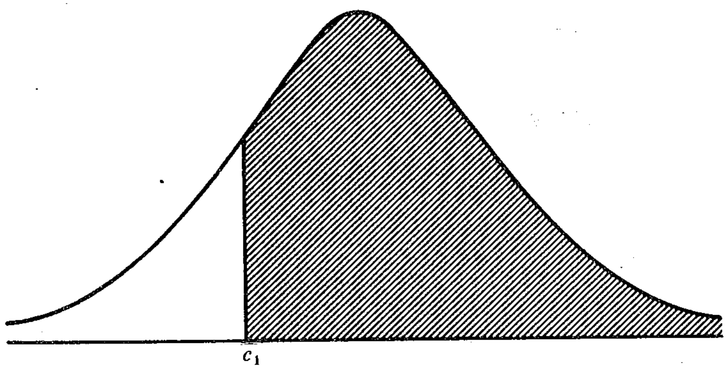

Figure A.1. Normal distribution truncated from below

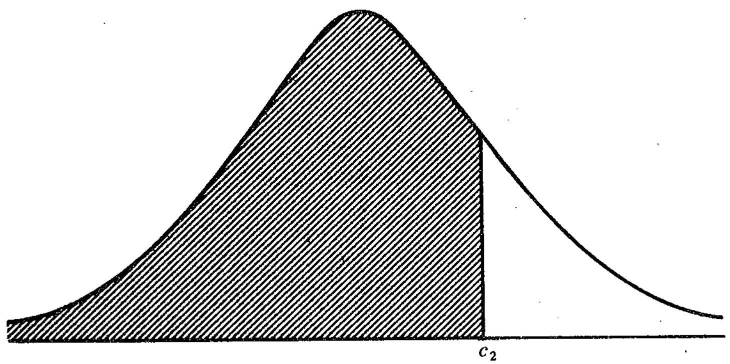

Figure A.2. Normal distribution truncated from above

If the distribution is doubly truncated, so that we consider $c_1 \leqslant X \leqslant c_2$, then

$$E(X) = \frac{\phi(c_1) - \phi(c_2)}{\Phi(c_2) - \Phi(c_1)} = M \quad \text{(say)}$$

$$V(X) = 1 - M^2 + \frac{c_1\phi(c_1) - c_2\phi(c_2)}{\Phi(c_2) - \Phi(c_1)}$$

If X has the normal distribution, with mean μ and variance σ^2 (instead of mean 0 and variance 1), then in the preceding formulas we have to substitute $(X-\mu)/\sigma$, $(c_1-\mu)/\sigma$, and $(c_2-\mu)/\sigma$ for X, c_1, and c_2 respectively (Johnson and Kotz, 1970, pp. 81–3).

In equations (5.46) and (5.47) in Chapter 5, we needed expressions of the form $E(u_1 \mid u_2 < c)$ and $E(u_1 \mid u_2 > c)$, where (u_1, u_2) are jointly normal, with zero means, unit variances, and covariance σ_{12}. Note that $E(u_1 \mid u_2) = \sigma_{12} u_2$. Thus,

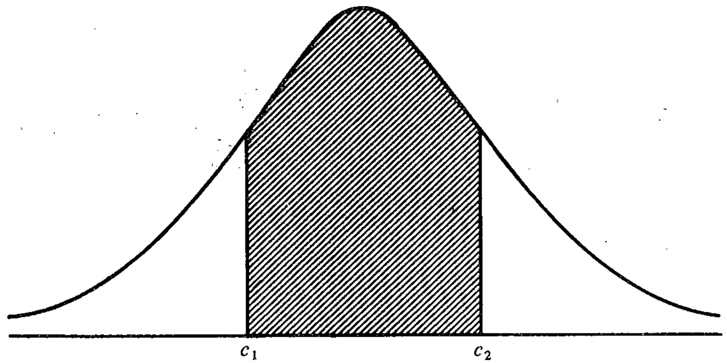

Figure A.3. Doubly truncated normal distribution

$$E(u_1 \mid u_2 < c) = \sigma_{12} E(u_2 \mid u_2 < c)$$

$$= -\sigma_{12} \frac{\phi(c)}{\Phi(c)}$$

Similarly,

$$E(u_1 \mid u_2 > c) = \sigma_{12} E(u_2 \mid u_2 > c)$$

$$= \sigma_{12} \frac{\phi(c)}{1 - \Phi(c)}$$

These are the expressions used in (5.46) and (5.47).

Derivatives

Very often we need the derivatives of functions of the form $\Phi(\alpha/\lambda)$ and $\phi(\alpha/\lambda)$ with respect to the parameters α and λ. Noting the expressions for Φ and ϕ, it can be verified (using the standard formulas for finding a derivative of an expression with an integral sign) that

$$\frac{\partial \Phi}{\partial \alpha} = \frac{1}{\lambda} \phi\left(\frac{\alpha}{\lambda}\right)$$

$$\frac{\partial \Phi}{\partial \lambda} = -\frac{\alpha}{\lambda^2} \phi\left(\frac{\alpha}{\lambda}\right)$$

$$\frac{\partial \phi}{\partial \alpha} = -\frac{\alpha}{\lambda^2}\phi\left(\frac{\alpha}{\lambda}\right)$$

$$\frac{\partial \phi}{\partial \lambda} = \frac{\alpha^2}{\lambda^3}\phi\left(\frac{\alpha}{\lambda}\right)$$

Moments of the truncated bivariate normal distribution

See the work of Rosenbaum (1961). Let

$$f(x,y,\rho) = \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left[-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right]$$

be the standard bivariate normal distribution, with zero means, unit variances, and correlation coefficient ρ.

Consider the truncated distribution $x \geqslant h,\ y \geqslant k$. Let

$$F(h,k,\rho) = \text{Prob}(x \geqslant h,\ y \geqslant k)$$

$$m_{ij} = E(x^i y^j \mid x \geqslant h,\ y \geqslant k)$$

$$h^* = \frac{h-\rho k}{(1-\rho^2)^{1/2}}$$

$$k^* = \frac{k-\rho h}{(1-\rho^2)^{1/2}}$$

$$Z = (h^2 - 2\rho hk + k^2)^{1/2}/(1-\rho^2)^{1/2}$$

Then, writing F for $F(h,k,\rho)$, we have

$$Fm_{10} = \phi(h)[1-\Phi(k^*)] + \rho\phi(k)[1-\Phi(h^*)]$$

$$Fm_{20} = F + h\phi(h)[1-\Phi(k^*)] + \rho^2 k\phi(k)[1-\Phi(h^*)] + \frac{\rho(1-\rho^2)^{1/2}}{(2\pi)^{1/2}}\phi(Z)$$

m_{01} and m_{02} can be obtained by interchanging h and k in these expressions.

$$Fm_{11} = F\rho + \rho h\phi(h)[1-\Phi(k^*)] + \rho k\phi(k)[1-\Phi(h^*)] + \frac{(1-\rho^2)^{1/2}}{(2\pi)^{1/2}}\phi(Z)$$

Note that in this example we considered F as $\text{Prob}(x \geqslant h,\ y \geqslant k)$.

Some nonnormal distributions

See the study of Lee (1982c, "Some Approaches to the Correction of Selectivity Bias," Appendix). Let $f(u)$ denote the density function and $F(u)$ the distribution function, so that $F(c) = \text{Prob}(u \leqslant c)$. Let

$$M_1(c) = E(u \mid u \leqslant c) \quad \text{and} \quad M_2(c) = E(u^2 \mid u \leqslant c)$$

Then we have the following results.

(1) *Normal distribution.*

$$\phi(u) = f(u) = (2\pi)^{-1/2}\exp(-\tfrac{1}{2}u^2) \quad (-\infty < u < \infty)$$

As defined earlier, we have

$$M_1(c) = -\phi(c)/\Phi(c), \quad M_2(c) = 1 - c\phi(c)/\Phi(c)$$

(2) *Log-normal distribution.* This would be useful to model variables with a skew distribution; $u = e^v$, where v is a standard normal variate.

$$M_1(c) = e^{1/2}\Phi(\ln c - 1)/F(c)$$
$$M_2(c) = e^2\Phi(\ln c - 2)/F(c)$$

(3) *Logistic distribution.*

$$f(u) = e^{-u}/(1+e^{-u})^2 \quad (-\infty < u < \infty)$$

$$M_1(c) = c + \ln[1 - F(c)]/F(c)$$

$$M_2(c) = \left[\frac{\pi^2}{3} + \text{sign}(c)\sum_{j=1}^{\infty}(-1)^{j-1}j^{-2}\Gamma_{j|c|}(3)\right] \Big/ F(c)$$

where $\Gamma_x(a) = \int_0^x u^{a-1}e^{-u}\,du$ is the incomplete gamma function with parameter a.

(4) *Uniform distribution.*

$$f(u) = 1 \quad (0 \leqslant u \leqslant 1)$$
$$M_1(c) = \tfrac{1}{2}c \qquad M_2(c) = c^2/3 \quad (0 \leqslant c \leqslant 1)$$

(5) *Laplace distribution.*

$$f(u) = \tfrac{1}{2}e^{-|u|} \quad -\infty < u < \infty$$
$$M_1(c) = c - 1 \quad \text{for } c \leqslant 0$$
$$M_1(c) = -(c+1)f(c)/F(c) \quad \text{for } c \geqslant 0$$
$$M_2(c) = c^2 - 2c + 2 \quad \text{for } c \leqslant 0$$
$$M_2(c) = [\tfrac{3}{2} - (c^2 + c + 1)f(c)]/F(c) \quad \text{for } c \geqslant 0$$

(6) *Exponential distribution.*

$$f(u) = (1/\sigma)e^{-u/\sigma} \quad (\sigma > 0,\ u > 0)$$